



Pengembangan Instrumen *Sound Wave Two Tier (Sowatt) Diagnostic Test: Analisis Rasch Model*

Muhammad Fadlan¹, Nour Fadillah¹, Winny Liliawati^{1*}, Ika Mustika Sari¹

¹Program Studi Pendidikan Fisika, Universitas Pendidikan Indonesia

*Corresponding author: winny@upi.edu

Article History:

Received: Juni 12, 2024

Revised: Juni 27, 2024

Accepted: Juli 03, 2024

Published: Desember 31, 2024

Keywords: *Diagnostic test, rasch model, sound wave*

Abstract: *This study aims to describe the validity of a sound wave diagnostic test instrument that can be used to identify students' conceptions. The research method used was Research and Development with the 4D model. The test instrument consisting of 15 questions with a two-tier format was tested on 64 students at one of the high schools in Bandung City. The results of the content validity of the instrument by experts were analyzed using a multi-rater test, namely the Many Facet Rasch Model (MRFM). As for construct validity, it was analyzed using Rasch modeling which describes validity, reliability, differentiating power, and item difficulty. The results showed that based on item dimensionality it was declared valid and every item met the criteria for item suitability with details of 3 easy questions, 11 medium questions, and 1 difficult question. Reliability based on Cronbach's alpha value is in the very good category, item reliability value is in the very good category, and person reliability is in the good category. The instrument can discriminate between individuals who have low and high abilities as seen from the results of the differentiating power test with all items in the very good category.*

Abstrak: Penelitian ini bertujuan untuk mendeskripsikan validitas instrumen tes diagnostik gelombang bunyi yang dapat digunakan untuk mengidentifikasi konsepsi siswa. Metode penelitian yang digunakan adalah *Research and Development* dengan model 4D. Instrumen tes yang terdiri dari 15 soal dengan format *two tier* diuji cobakan kepada 64 orang siswa pada salah satu SMA di Kota Bandung. Hasil validitas isi instrumen oleh ahli dianalisis menggunakan uji *multi rater*, yaitu *Many Facet Rasch Model (MRFM)*. Sedangkan untuk validitas konstruk dianalisis menggunakan pemodelan Rasch yang menjabarkan validitas, reliabilitas, daya pembeda, dan tingkat kesulitan butir soal. Hasil penelitian menunjukkan bahwa berdasarkan *item dimensionality* dinyatakan valid dan setiap butir soal memenuhi kriteria kesesuaian butir soal dengan rincian 3 soal mudah, 11 soal sedang, dan 1 soal sulit. Reliabilitas berdasarkan nilai Cronbach's *alpha* termasuk kategori tinggi, nilai *item reliability* termasuk kategori sangat baik, dan *person reliability* termasuk kategori baik. Instrumen dapat membedakan individu yang memiliki kemampuan rendah dan tinggi yang terlihat dari hasil uji daya pembeda dengan setiap butir soal termasuk kategori sangat baik. Kesimpulannya, butir soal pada instrumen tes diagnostik gelombang bunyi yang dikembangkan melalui model 4D dengan memanfaatkan analisis pemodelan Rasch dapat digunakan untuk mengukur tingkat konsepsi siswa SMA.

PENDAHULUAN

Pembelajaran fisika di tingkat sekolah menengah memiliki tuntutan yang ideal. Tuntutan tersebut meliputi penguasaan konsep fisika serta pengembangan keterampilan berpikir

kritis, analitis, dan kreatif. Menurut Arends (2012), tujuan pembelajaran fisika pada level sekolah menengah adalah agar siswa memiliki pemahaman yang mendalam tentang prinsip-prinsip fisika dan mampu mengaplikasikannya dalam

berbagai konteks kehidupan sehari-hari. Berdasarkan hal tersebut, maka dalam proses pembelajaran fisika, siswa diarahkan untuk dapat memahami konsep dan materi pembelajaran dengan sebaik-baiknya.

Faktanya saat proses pembelajaran berlangsung, siswa tidak selalu memahami konsep secara utuh, terlebih lagi pada mata pelajaran fisika yang memuat banyak konsep dan seringkali siswa tidak menyadari bahwa dirinya tidak memahami konsep secara utuh (Syahrul, 2015). Pemahaman konsep yang tidak utuh disebut juga sebagai miskonsepsi. Miskonsepsi diartikan sebagai suatu kesalahpahaman atas suatu konsep yang sudah disepakati para ahli (Ormrod, 2009). Berdasarkan hasil temuan beberapa peneliti dari berbagai materi pembelajaran fisika, gelombang bunyi menjadi salah satu materi yang dalam pembelajarannya siswa masih mengalami miskonsepsi (Hrepic, et al., 2010; Nofriati et al., 2016; Widiastuti et al., 2019; Gani et al., 2023).

Salah satu cara untuk mendiagnosis tingkat konsepsi siswa adalah dengan melakukan tes diagnostik (Anggraeni, 2017). Tes diagnostik dapat digunakan oleh guru untuk mengetahui konsepsi awal siswa dan kesulitan siswa dalam pembelajaran agar dapat dilakukan pemilihan metode pembelajaran yang sesuai untuk mengatasi kesulitan tersebut (Kaltakçi & Didiş, 2007; Gurel et al., 2015; Ayar et al., 2015). Bentuk tes diagnostik yang dapat digunakan adalah tes dua tingkat berbentuk pilihan ganda. Bentuk soal tersebut dipilih karena selain siswa mengerjakan butir soal yang mengungkapkan konsep tertentu, siswa juga harus mengungkapkan alasan memilih jawaban tersebut serta dapat digunakan untuk membantu menguji pemahaman siswa serta membantu mengidentifikasi tingkat konsepsi yang dimiliki oleh siswa (Suwanto, 2013; Nofiana et al., 2014; Julianda et al., 2022).

Sebelum disebarluaskan, instrumen tes diagnostik perlu diuji kelayakannya agar memiliki validitas yang baik dan dapat digunakan secara tepat untuk mengukur tingkat pemahaman konsep siswa. Selain itu, instrumen yang layak juga memberikan hasil analisis diagnostik yang dapat dipercaya. Pada akhirnya, hasil evaluasi yang diberikan dari instrumen tes diagnostik yang dikembangkan dapat dengan jelas menggambarkan tingkat pemahaman konsep siswa (Maulana et al., 2023).

Pengujian kelayakan instrumen tes diagnostik, digunakan analisis pemodelan Rasch. Analisis Rasch meliputi uji parameter siswa dan soal yang diujikan (Sumintono & Widhiarso, 2015; Gani et al., 2023). Selain itu, Rasch model juga dapat digunakan dalam menentukan indeks reliabilitas hingga analisis item ke masing-masing tingkat, reliabilitas responden, unidimensionalitas, hingga menentukan kemampuan siswa dengan kemungkinan seorang siswa menjawab suatu item dengan benar akan meningkat secara monoton dengan kemampuan yang dimiliki siswa tersebut (Suryani, 2018; Xiao et al., 2018; Islam et al., 2020; Maulana et al., 2023).

Analisis Rasch Model menggunakan pendekatan probabilitas dalam melihat objek pengukuran sehingga analisis Rasch Model tidak deterministik dan dapat mengidentifikasi objek yang diukur lebih akurat serta dapat mendeskripsikan interaksi antara responden dengan butir soal (Sumintono & Widhiarso, 2015; Putra et al., 2021; Indihadi et al., 2022; Salsabila et al., 2023).

Penelitian ini dikembangkan instrumen diagnostik *Sound Wave Two Tier (SOWATT) Diagnostic Test* berbentuk pilihan ganda beralasan (*two-tier question*) yang dapat digunakan untuk mengidentifikasi konsepsi siswa pada materi gelombang bunyi melalui implementasi dan analisis Rasch model.

METODE PENELITIAN

Penelitian pengembangan instrumen SOWATT *Diagnostic Test* menggunakan metode *Research and Development* (R&D). Desain penelitian yang digunakan adalah model 4D yang dikembangkan oleh Thiagarajan pada tahun 1974. Tahapan pada model 4D ditunjukkan pada Gambar 1 berikut.



Gambar 1. Tahapan Model 4D

Instrumen Penelitian

Instrumen tes diagnostik yang dinamakan SOWATT terdiri dari 15 butir soal dengan format *two tier*. *Tier* pertama berupa pertanyaan dalam bentuk pilihan ganda dan *tier* kedua berupa pilihan alasan terhadap *tier* pertama. Instrumen yang digunakan untuk mengidentifikasi tingkat konsepsi siswa ini disebarakan kepada 64 siswa SMA kelas XI pada salah satu SMA di Kota Bandung.

Analisis Data

Instrumen SOWATT dianalisis berdasarkan validitas isi dan validitas konstruk. Validitas isi diperoleh dari hasil penilaian ahli terhadap instrumen SOWATT. Validator terdiri dari enam orang ahli dengan rincian tiga orang dosen

fisika dan tiga orang guru mata pelajaran fisika. Hasil penilaian tersebut dianalisis dengan uji *multi-rater*, yaitu *Many Facet Rasch Model* (MRFM) menggunakan *software* Minifac versi 3.85.1. Sedangkan validitas konstruk diperoleh dari respon jawaban siswa yang telah mengerjakan instrumen SOWATT yang diolah menjadi skor dan dianalisis dengan pemodelan Rasch menggunakan *software* Winsteps versi 4.5.0. Instrumen dianalisis berdasarkan hasil *Wright map*, *item dimensionality*, kesesuaian butir soal, nilai *Cronbach's alpha*, nilai *item and person reliability*, uji daya pembeda, dan tingkat kesulitan butir soal. Informasi pertama yang dianalisis pada pemodelan Rasch adalah *Wright map* yang menggambarkan sebaran kemampuan siswa dan tingkat kesulitan soal dengan skala yang sama (Sumintono & Widhiarso, 2015).

Instrumen dapat dikatakan valid apabila 1) nilai *raw variance explained by measure* lebih dari 20%, dan 2) nilai *unexplained variance in 1st contrast* untuk Eigenvalue kurang dari 3 dan nilai *observed* kurang dari 15 (Fisher, 2007; Sumintono & Widhiarso, 2015; Adams et al., 2018). Kriteria untuk hasil dari setiap analisis model Rasch ditunjukkan pada Tabel 1 (Boone et al., 2014; Smiley, 2015; Sumintono & Widhiarso, 2015).

Tabel 1. Kriteria Hasil Analisis Rasch Model

Kesesuaian Butir Soal	
Kriteria	Nilai
<i>Outfit</i> MNSQ	0,50 < MNSQ < 1,50
<i>Outfit</i> ZSTD	-2,00 < ZSTD < 2,00
<i>Pt Measure Corr</i>	0,40 < <i>Pt Me Corr.</i> < 0,85
Cronbach's alpha (KR-20)	
Nilai Indeks	Interpretasi
$KR - 20 \geq 0,80$	Sangat tinggi
$0,70 \leq KR - 20 < 0,80$	Tinggi
$0,60 \leq KR - 20 < 0,70$	Baik
$0,50 \leq KR - 20 < 0,60$	Sedang
$KR - 20 < 0,50$	Rendah
Item and Person reliability	
Nilai Indeks	Interpretasi
$r > 0,94$	Istimewa
$0,90 < r \leq 0,94$	Sangat baik
$0,80 < r \leq 0,90$	Baik
$0,67 < r \leq 0,80$	Cukup
$r \leq 0,67$	Rendah

Daya Pembeda	
Pt Measure Corr.	Interpretasi
$ID \geq 0,40$	Sangat baik
$0,40 < ID \leq 0,30$	Baik
$0,30 < ID \leq 0,20$	Kurang baik
$ID < 0,20$	Buruk
Tingkat Kesulitan Butir Soal	
Kriteria	Interpretasi
$ME < -1SD$	Mudah
$-1SD \leq ME \leq +1SD$	Sedang
$ME > +1SD$	Sulit

HASIL DAN PEMBAHASAN

Hasil dan pembahasan meliputi tahapan pengembangan instrumen SOWATT menggunakan model 4D.

Tahap Define

Pada tahap *define*, dilakukan studi literatur untuk mengetahui instrumen tes diagnostik berformat *two tier* pada materi gelombang bunyi pada penelitian-penelitian terdahulu. Studi literatur juga dilakukan untuk menentukan konsep dari gelombang bunyi yang digunakan dalam penyusunan instrumen SOWATT. Konsep dari materi gelombang bunyi yang digunakan dalam penelitian ini adalah konsep karakteristik gelombang bunyi, resonansi bunyi, dan efek Doppler.

Tahap Design

Pada tahap *design*, dilakukan perancangan sebaran kisi-kisi soal. Sebaran kisi-kisi soal terdiri dari konsep, indikator pencapaian kompetensi (IPK), dan nomor soal untuk setiap IPK. Kemudian konten soal disusun sesuai dengan sebaran kisi-kisi yang telah dibuat. Tabel 2 menunjukkan jumlah soal untuk setiap indikator pencapaian kompetensi.

Tabel 2. Sebaran Kisi-Kisi Soal

Konsep	IPK	Nomor Soal
Karakteristik Gelombang Bunyi	Membedakan frekuensi dan amplitudo gelombang bunyi	2
	Menganalisis cepat rambat bunyi di berbagai medium	1, 5
	Menganalisis perambatan gelombang	6

Konsep	IPK	Nomor Soal
	bunyi pada medium udara	3
	Menganalisis cepat rambat bunyi apabila dipengaruhi oleh medium, frekuensi, dan amplitudo gelombang bunyi.	
	Memprediksi cepat rambat bunyi apabila pendengar memiliki jarak yang sama terhadap sumber bunyi	4, 7
	Resonansi Bunyi	Menganalisis hubungan panjang senar terhadap nada bunyi yang dihasilkan pada fenomena dawai
	Menganalisis pola gelombang yang terbentuk pada dawai, pipa organa terbuka, atau pipa organa tertutup	9, 10
	Membandingkan frekuensi nada dasar yang dihasilkan pipa organa terbuka dan pipa organa tertutup	11
	Merumuskan persamaan frekuensi yang dihasilkan pada resonansi harmonik tertentu	13
	Menganalisis hubungan antara panjang kolom udara dengan bunyi yang dihasilkan pada fenomena resonansi	12
Efek Doppler	Menganalisis frekuensi dan panjang gelombang pada peristiwa efek Doppler	14
	Memprediksi perubahan intensitas atau frekuensi pada peristiwa efek Doppler	15

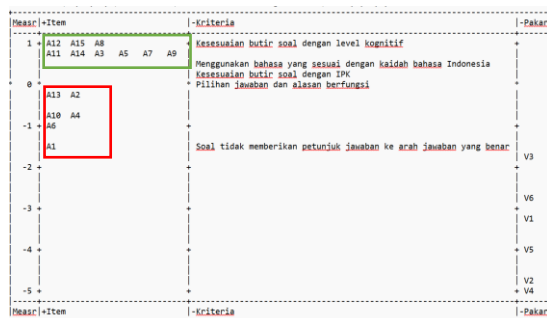
Tahap Develop

Pada tahap *develop*, instrumen SOWATT diagnostic test dengan format *two tier*. Gambar 2 menunjukkan salah satu butir soal yang terdapat pada instrumen SOWATT.

- 1.1 Seruling merupakan salah satu alat musik yang memanfaatkan kolom udara sebagai sumber bunyi dengan bentuk pipa organa terbuka. Jika seruling ditiup, maka akan terdengar bunyi tiupan yang menyebabkan udara dalam kolom udara bergetar. Berdasarkan hal tersebut, seruling dapat menghasilkan resonansi harmonik. Bagaimanakah persamaan untuk mencari frekuensi pada resonansi harmonik ketiga?
- $f = \frac{v}{l}$
 - $f = \frac{v}{2l}$
 - $f = \frac{v}{3l}$
 - $f = \frac{3v}{2l}$
 - $f = \frac{3v}{3l}$
- 1.2 Alasan memilih jawaban 1.1:
- Resonansi harmonik ketiga sama dengan nada atas pertama dan terbentuk ketika $l = \frac{\lambda}{2}$.
 - Resonansi harmonik ketiga sama dengan nada atas pertama dan terbentuk ketika $l = \lambda$.
 - Resonansi harmonik ketiga sama dengan nada atas kedua dan terbentuk ketika $l = \lambda$.
 - Resonansi harmonik ketiga sama dengan nada atas kedua dan terbentuk ketika $l = \frac{3\lambda}{2}$.
 - Resonansi harmonik ketiga sama dengan nada atas ketiga dan terbentuk ketika $l = \frac{3\lambda}{2}$.

Gambar 2. Butir Soal Instrumen SOWATT

Selanjutnya instrumen SOWATT diuji validitas isinya oleh enam ahli dengan kriteria kesesuaian butir soal dengan IPK; kesesuaian butir soal dengan level kognitif; pilihan jawaban dan alasan berfungsi, homogen serta logis dari segi materi; menggunakan bahasa yang sesuai dengan kaidah bahasa Indonesia; dan soal tidak memberikan petunjuk jawaban ke arah jawaban yang benar. Analisis validitas isi dilakukan dengan uji multi-rater, yaitu *Many Facet Rasch Model (MRFM)*.



Gambar 3. Hasil Uji Validitas Isi oleh Ahli

Berdasarkan Gambar 3 dapat terlihat bahwa terdapat empat kolom dengan kolom pertama menunjukkan nilai logit pengukuran yang berkisar pada rentang -5 hingga +1, kolom kedua menunjukkan butir soal yang dikodekan dengan huruf A dan nomor urut butir soal (A1–A15), kolom ketiga menunjukkan kriteria penilaian instrumen, dan kolom keempat menunjukkan validator yang dikodekan dengan huruf V dan nomor urut validator (V1–V6).

Berdasarkan kolom kedua, butir soal instrumen SOWATT berada pada rentang nilai logit terkecil mendekati -2 hingga nilai terbesar mendekati +1. Butir soal yang berada pada nilai logit positif (>0) dikategorikan sebagai soal yang baik, sedangkan butir soal dengan nilai logit negatif (< 0) dikategorikan sebagai soal yang kurang baik (Darmana et al., 2021). Sehingga dapat disimpulkan bahwa butir soal pada kotak berwarna hijau merupakan butir soal yang baik, sedangkan butir soal pada kotak berwarna merah merupakan butir soal yang kurang baik sehingga perlu diperbaiki. Butir soal yang kurang baik tersebut hanya memenuhi dua kriteria yaitu pilihan jawaban dan alasan berfungsi, dan soal tidak memberikan petunjuk jawaban ke arah jawaban yang benar. Maka, harus diperbaiki kembali sehingga memenuhi semua kriteria. Adapun yang perlu diperbaiki yaitu penyesuaian kembali butir soal dengan level kognitif, penyesuaian kembali bahasa sesuai kaidah bahasa Indonesia, dan penyesuaian kembali butir soal dengan IPK.

Selain itu, hasil uji validitas isi juga dapat terlihat dari nilai *exact agreements* dan nilai *expected agreements* yang ditunjukkan Tabel 3 berikut.

Tabel 3. Perbandingan Nilai *Exact Agreements* dan *Expected Agreements*

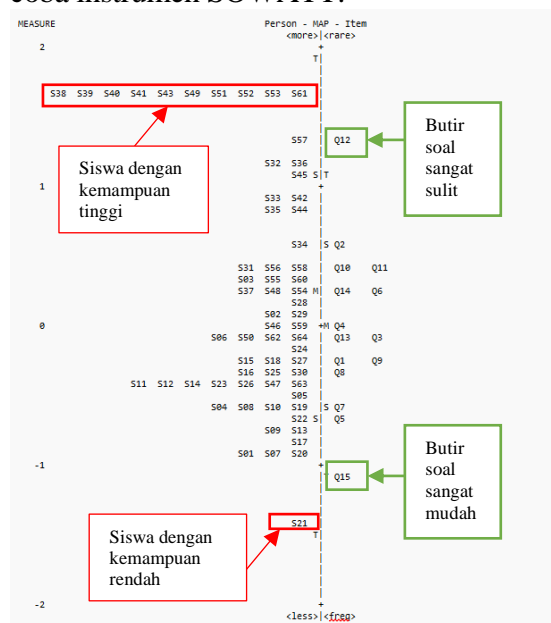
<i>Exact Agreements</i>	<i>Expected Agreements</i>
89,60%	90,10%

Berdasarkan Tabel 3 dapat terlihat bahwa nilai *exact agreements* sebesar 89,60% dan nilai *expected agreements* sebesar 90,10%. Angka tersebut mengindikasikan bahwa keenam ahli memiliki penilaian yang hampir sama terhadap instrumen SOWATT. Berdasarkan hasil tersebut, instrumen SOWATT *diagnostic test* layak untuk digunakan pada uji coba lapangan.

Tahap Disseminate

Pada tahap *disseminate*, instrumen SOWATT yang telah dinilai oleh ahli, kemudian direvisi berdasarkan masukan dari

keenam ahli. Selanjutnya instrumen SOWATT disebarakan untuk diuji cobakan kepada siswa. Hasil analisis pertama ditunjukkan oleh Wright map yang menggambarkan sebaran kemampuan siswa dan sebaran tingkat kesulitan soal dengan skala yang sama. Gambar 4 menunjukkan Wright map yang didapatkan dari data uji coba instrumen SOWATT.



Gambar 4. Wright Map Siswa dan Butir Soal

Gambar 4 menunjukkan sebaran kemampuan siswa dalam menjawab butir soal dan tingkat kesulitan butir soal. *Wright map* sebelah kiri menggambarkan kemampuan siswa dan terlihat bahwa siswa dengan kode S38, S39, S40, S41, S43, S49, S51, S52, S53, dan S61 (10 orang) memiliki kemampuan yang paling tinggi dalam mengerjakan soal pada instrumen SOWATT dibandingkan dengan siswa lainnya. Sedangkan untuk siswa yang memiliki kemampuan paling rendah dalam mengerjakan soal adalah siswa dengan kode S21 (1 orang). Sedangkan untuk sebaran tingkat kesulitan butir soal, butir soal dengan kode Q12 merupakan butir soal yang termasuk dalam kategori sangat sulit serta butir soal dengan kode Q15 merupakan butir soal yang termasuk dalam kategori sangat mudah.

Hasil dari validitas dan reliabilitas butir soal dijelaskan sebagai berikut. Unidimensionalitas instrumen ditunjukkan Tabel 4 berikut.

Tabel 4. Unidimensionalitas Instrumen

Raw variance explained by measure	Unexplained variance in 1 st contrast	
	Eigenvalue	Observed
49,19%	3,17	10,60%

Tabel 4 menunjukkan bahwa nilai *raw variance explained by measures* adalah 49,90%. Menurut Sumintono & Widhiarso (2015), apabila nilai *raw variance explained by measures* lebih dari 20%, maka instrumen yang digunakan dapat mengukur satu variabel tanpa dipengaruhi variabel-variabel yang lain. Adapun untuk nilai *unexplained variance in 1st contrast Eigenvalue* sebesar 3,17 (berada pada rentang 3) dan nilai *observed* sebesar 10,60% (kurang dari 15%). Berdasarkan hasil tersebut, instrumen SOWATT dinyatakan sebagai instrumen yang valid. Kesesuaian setiap butir soal, dapat terlihat dari nilai *Outfit* MNSQ dan ZSTD serta nilai *Pt Mean Corr.* yang ditunjukkan Tabel 5.

Tabel 5. Kesesuaian Butir Soal

Nomor Butir Soal	Nilai Outfit		Pt Measure Corr.
	MNSQ	ZSTD	
1	0,95	-0,07	0,63
2	1,13	0,64	0,43
3	1,24	0,88	0,62
4	1,22	0,85	0,57
5	0,92	-0,10	0,54
6	0,90	-0,30	0,68
7	0,73	-0,72	0,52
8	0,61	-1,38	0,70
9	0,74	-0,85	0,69
10	0,69	-1,18	0,80
11	1,45	1,53	0,62
12	1,01	1,74	0,44
13	0,95	-0,10	0,66
14	0,68	-1,27	0,76
15	0,97	0,09	0,41

Berdasarkan Tabel 5 dapat terlihat bahwa setiap butir soal memenuhi ketiga kriteria nilai *Outfit* MNSQ, nilai *Outfit* ZSTD, dan *Pt Mean Corr.* Menurut Sumintono & Widhiarso (2015), apabila butir soal memenuhi ketiga kriteria tersebut, maka butir-butir soal instrumen SOWATT telah layak digunakan untuk mengukur

tingkat konsepsi siswa pada materi gelombang bunyi. Sedangkan untuk nilai dari *Cronbach's alpha*, *item reliability*, dan *person reliability* ditunjukkan pada Gambar 5.

The image shows two SPSS output tables. The top table is titled 'SUMMARY OF 64 MEASURED Person' and lists various statistics including Mean (24.9), SEM (1.4), P.S.D (11.4), S.SD (11.5), MAX. (42.0), MIN. (4.0), and reliability values: REAL RMSE (.32), MODEL RMSE (.31), S.E. OF Person MEAN (.11), Person RELIABILITY (.85), and CRONBACH ALPHA (KR-20) Person RAW SCORE 'TEST' RELIABILITY (.88). The bottom table is titled 'SUMMARY OF 15 MEASURED Item' and lists statistics for 15 items, including Mean (106.1), SEM (8.3), P.S.D (31.1), S.SD (32.2), MAX. (164.0), MIN. (39.0), and reliability values: REAL RMSE (.14), MODEL RMSE (.13), S.E. OF Item MEAN (.15), Item RELIABILITY (.94), and CRONBACH ALPHA (KR-20) Item RAW SCORE 'TEST' RELIABILITY (.94).

Gambar 5. Hasil dari *Cronbach's Alpha*, *Item Reliability*, dan *Person Reliability*

Berdasarkan Gambar 5 dapat terlihat bahwa *person reliability* bernilai 0,85 dengan interpretasi baik dan *item reliability* bernilai 0,94 dengan interpretasi sangat baik. Hal ini menunjukkan bahwa konsistensi jawaban dari siswa baik serta kualitas instrumen SOWATT dinyatakan sangat baik dalam hal konsistensi bobot soal dan pilihan jawabannya. Kemudian untuk nilai *Cronbach's alpha* (KR-20) sebesar 0,88 dengan interpretasi tinggi. Hal ini menunjukkan bahwa instrumen SOWATT memiliki kualitas yang sangat baik karena dapat mengidentifikasi hubungan antara siswa (*person reliability*) dengan butir soal (*item reliability*). Berdasarkan hasil analisis tersebut dapat disimpulkan bahwa instrumen SOWATT reliabel untuk digunakan.

Analisis berikutnya adalah uji daya pembeda setiap butir soal yang digunakan untuk menunjukkan seberapa jauh sebuah butir soal mampu membedakan individu yang memiliki kemampuan yang tinggi dan rendah. Hasil uji daya pembeda ditunjukkan Tabel 6 berikut.

Tabel 6. Hasil Uji Daya Pembeda

Butir Soal	Nilai Pt Mean Corr	Interpretasi
1	0,63	Sangat baik
2	0,43	Sangat baik
3	0,62	Sangat baik
4	0,57	Sangat baik
5	0,54	Sangat baik
6	0,68	Sangat baik
7	0,52	Sangat baik
8	0,70	Sangat baik
9	0,69	Sangat baik
10	0,80	Sangat baik
11	0,62	Sangat baik
12	0,44	Sangat baik
13	0,66	Sangat baik
14	0,76	Sangat baik
15	0,41	Sangat baik

Berdasarkan Tabel 6 dapat terlihat bahwa 15 butir soal (100%) dinyatakan memiliki daya pembeda yang sangat baik. Interpretasi ini menunjukkan bahwa butir soal pada instrumen SOWATT mampu membedakan individu yang memiliki kemampuan yang tinggi dan rendah. Analisis berikutnya adalah tingkat kesulitan setiap butir soal dengan membandingkan nilai *logit measure* (ME) dengan *standar deviasi* (SD). Nilai SD yang diperoleh adalah +0,55, sehingga dapat dibuat kriteria yang ditunjukkan Tabel 7 berikut.

Tabel 7. Hasil Interpretasi Tingkat Kesulitan Butir Soal

Kriteria	Interpretasi
ME < -0,55	Mudah
-0,55 ≤ ME ≤ +0,55	Sedang
ME > +0,55	Sulit

Kemudian tingkat kesulitan setiap butir soal diinterpretasikan yang ditunjukkan pada Tabel 8 berikut.

Tabel 8. Hasil Interpretasi Tingkat Kesulitan Butir Soal

Nomor Butir Soal	Measure (ME)	Kriteria	Interpretasi
1	-0,23	-0,55 ≤ ME ≤ 0,55	Sedang
2	0,55	-0,55 ≤ ME ≤ 0,55	Sedang
3	-0,06	-0,55 ≤ ME ≤ 0,55	Sedang
4	0,02	-0,55 ≤ ME ≤ 0,55	Sedang
5	-0,65	ME < -0,55	Mudah
6	0,21	-0,55 ≤ ME ≤ 0,55	Sedang
7	-0,56	ME < -0,55	Mudah
8	-0,29	-0,55 ≤ ME ≤ 0,55	Sedang
9	-0,28	-0,55 ≤ ME ≤ 0,55	Sedang
10	0,45	-0,55 ≤ ME ≤ 0,55	Sedang
11	0,44	-0,55 ≤ ME ≤ 0,55	Sedang
12	1,30	ME > 0,55	Sulit
13	-0,12	-0,55 ≤ ME ≤ 0,55	Sedang
14	0,26	-0,55 ≤ ME ≤ 0,55	Sedang
15	-1,06	ME < -0,55	Mudah

Berdasarkan Tabel 8 dapat terlihat bahwa butir soal nomor 5, 7, dan 15 termasuk kategori mudah; butir soal nomor 1, 2, 3, 4, 6, 8, 9, 10, 11, 13, dan 14 termasuk kategori sedang; dan butir soal nomor 12 termasuk kategori sulit pada konsep efek Doppler.

SIMPULAN

Penelitian ini menyimpulkan bahwa butir soal *Sound Wave Two Tier (SOWATT) Diagnostic Test* dapat digunakan untuk mengukur pemahaman konsep siswa SMA pada materi gelombang bunyi. Nilai dari *Cronbach's alpha*, *item reliability*, dan *person reliability* menunjukkan bahwa instrumen SOWATT memiliki kualitas yang baik karena dapat mengidentifikasi hubungan antara person dan item atau konsistensi yang baik dari jawaban siswa terhadap butir soal. Berdasarkan uji daya pembeda, butir soal SOWATT mampu membedakan siswa yang memiliki kemampuan yang tinggi dan rendah. Tingkat kesulitan butir soal telah mewakili untuk setiap kategori yang meliputi mudah, sedang dan sulit. Dengan demikian, instrumen SOWATT yang dianalisis menggunakan pemodelan Rasch dinyatakan valid dan reliabel untuk diujicobakan pada skala yang lebih luas.

DAFTAR PUSTAKA

- Adams, D., Sumintono, B., Mohamed, A., & Noor, N. S. M. (2018). E-learning readiness among students of diverse backgrounds in a leading Malaysian higher education institution. *Malaysian Journal of Learning and Instruction*, 15(2), 227-256.
<https://doi.org/10.32890/mjli2018.15.2.9>
- Anggraeni, H. B., Bambang S., & Rizka, A. P. (2017). Pengembangan tes formatif yang berfungsi sebagai tes diagnostik kesulitan belajar pokok bahasan animalia. *Jurnal Prodi Pendidikan Biologi*, 6(6), 341-350.
- <https://doi.org/10.21831/edubio.v6i6.8131>
- Arends, R. (2012). *Learning to teach: tenth edition*. New York: McGraw-HillMcGraw-Hill Education.
- Ayar, M. C., Aydeniz, M., & Yalvac, B. (2015). Analyzing science activities in force and motion motion concepts: A design of an immersion unit. *International Journal of Science and Mathematics Education*, 13(1), 95-121.
<https://doi.org/10.1007/s10763-013-9476-y>
- Boone, W.J., Staver, J.R., & Yale, M.S. (2014). *Rasch Analysis in the human sciences*. Dordrecht: Springer.
- Darmana, A., Sutiani, A., Nasution, H. A., Sylvia, N. S. A., Aminah, N., & Utami, T. (2021). Analysis of multi-rater with facets on instruments HOTS of solution chemistry based on tawheed. *Journal of physics: conference series*, 1819(1).
<https://doi.org/10.1088/1742-6596/1819/1/012038>
- Fisher, W. P. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transactions*, 21(1), 1095.
<http://www.rasch.org/rmt/rmt211a.htm>
- Gani, A. W., Samsudin, A., Aviyanti, L., Aminudin, A. H., Kurniawan, F., Zahran, M., Kunaedi, J., & Kapici, H. O. (2023). Identifying students' conceptions of sound waves via the sound wave four-tier (SOFT) diagnostic test. *Berkala ilmiah pendidikan fisika*, 11(3), 318-328.
<http://dx.doi.org/10.20527/bipf.v11i3.16903>
- Gurel, D. K., Eryılmaz, A., & McDermott, L. C. (2015). A review and comparison of diagnostic instruments to identify students'

- misconceptions in science. *Eurasia Journal of Mathematics, Science & Technology Education*, 11(5), 989–1008.
<https://doi.org/10.12973/eurasia.2015.1369a>
- Hrepic, Z., Zollman, D.A., & Rebello, N.S. (2010). Identifying students' mental models of sound propagation: the role of conceptual blending in understanding conceptual change. *Physics Education Research*, 6(2), 1–18.
<https://doi.org/10.1103/PhysRevSTPER.6.020114>
- Indihadi, D., Suryana, D., & Ahmad, A. B. (2022). The analysis of construct validity of Indonesian creativity scale using Rasch model. *Creativity Studies*, 15(2), 560–576.
<https://doi.org/10.3846/cs.2022.15182>
- Islam, A. Y. M. A., Gu, X., Crook, C., & Spector, J. M. (2020). Assessment of ICT in tertiary education applying structural equation modeling and Rasch model. *SAGE Open*, 10(4).
<https://doi.org/10.1177/2158244020975409>
- Julianda, R., Saminan, A. Halim. (2022). Analisis miskonsepsi siswa dengan two tier diagnostic test pada materi gerak lurus di SMA negeri 3 Banda Aceh. *Journal of Technology and Literacy in Education*, 1(1), 14–20.
- Kaltakçı, D., & Didiş, N. (2007). Identification of pre-service physics teachers' misconceptions on gravity concept: a study with a 3-tier misconception test. *AIP Conference Proceedings*, 899(1), 499–500. AIP Publishing.
<https://doi.org/10.1063/1.2733255>
- Maulana, S., Rusilowati, A., Nugroho, S. E., & Susilaningsih, E. (2023). Implementasi Rasch model dalam pengembangan instrumen tes diagnostik. *Prosiding Seminar Nasional Pascasarjana*, 6(1), 748–757.
- Nofiana, N., Sajidan, S., & Puguh, P. (2014). Pengembangan instrumen evaluasi two-tier multiple choice question untuk mengukur keterampilan berpikir tingkat tinggi pada materi kingdom plantae. *Jurnal Inkuiri*, 3(11), 60–74.
<https://doi.org/10.20961/inkuiri.v3i2.9694>
- Nofriati, Kusairi, S. & Rahayu, S. (2016). Penguasaan konsep siswa SMP pada materi bunyi. *Prosiding semnas pendidikan IPA pascasarjana UM*, 1102–1111.
- Ormrod, J. E. (2009). *Psikologi pendidikan membantu siswa tumbuh dan berkembang: Jilid 1*. Jakarta: Erlangga.
- Putra, Z. H., Hermita, N., & Alim, J. A. (2021). Analisis pengetahuan matematika, didaktika, dan teknologi calon guru sekolah dasar menggunakan Rasch model. *Mosharafa: Jurnal Pendidikan Matematika*, 10(3), 345–356.
<https://karya.brin.go.id/id/eprint/15618>
- Salsabila, F., Nurihsan, J., Sunarya, Y. (2023). Pengujian validitas dan reliabilitas instrumen manajemen diri remaja: Rasch model analysis. *Jurnal bimbingan dan konseling terapan*, 7(1), 15–25.
<http://dx.doi.org/10.30598/jbkt.v7i1.1741>
- Smiley, J. (2015). Classical test theory or Rasch: a personal account from a novice user. *Shiken*, 19(1), 16–29.
- Sumintono & Widhiarso. (2015). *Aplikasi pemodelan Rasch pada asesmen pendidikan*. Cimahi: Trim Komunikata.
- Suryani, Y. E. (2018). Aplikasi Rasch model dalam mengevaluasi intelligenz structure test. *Psikohumaniora: jurnal penelitian*

- psikologi*, 3(1), 73-100.
<https://doi.org/10.21580/pjpp.v3i1.2052>
- Suwarto. (2013). *Pengembangan tes diagnostik dalam pembelajaran*. Yogyakarta: Pustaka Pelajar.
- Syahrul, D. (2015). Identifikasi miskonsepsi dan penyebab miskonsepsi siswa dengan three-tier diagnostic test pada materi dinamika rotasi. *Jurnal inovasi pendidikan fisika*, 4(3), 67–70.
<https://doi.org/10.26740/ipf.v4n3.p%25p>
- Widiastuti, W., Ari, S., & Purwanto, J. (2019). Remediasi miskonsepsi pada materi gelombang bunyi dengan pendekatan konstruktivisme metode 5E di SMAN 1 Turi. *Prosiding Seminar Nasional Fisika dan Aplikasinya*.
<https://jurnal.uns.ac.id/prosidingsnfa/article/view/35909/24936>.
- Xiao, Y., Han, J., Koenig, K., Xiong, J., & Bao, L. (2018). Multilevel rasch modeling of two-tier multiple-choice test: a case study using lawson's classroom test of scientific reasoning. *Physical review physics education research*, 14(2), 20104.
<https://doi.org/10.1103/PhysRevPhysicsEducRes.14.020104>